Total Variance, an Estimator of Long-Term Frequency Stability

C. A. Greenhall* D. A. Howe† D. B. Percival‡

August 11, 1998

Abstract

Total variance is a statistical tool developed for improved estimates of frequency stability at averaging times up to half the test duration. As a descriptive statistic, Total variance performs an exact decomposition of the sample variance of the frequency residuals into components associated with descending frequency octaves. As an estimator of Allan variance, Total variance has modest bias and has greater degrees of freedom and lesser mean square error than the standard estimator does.

1 Introduction and conclusions

Almost by definition, there can never be enough data when making long-term stability measurements of clocks and frequency standards. Having collected data during a time period T, we have to accept a tradeoff between averaging time τ and confidence in the estimate $\hat{\sigma}_y(\tau, T)$ of Allan deviation $\sigma_y(\tau)$. To improve this tradeoff, Howe, Allan, and Barnes [1] introduced the practice of incorporating all the available overlapping samples of the increment of τ-average frequency into the estimate. Of course, for the largest averaging time $\tau = T/2$ there is only one such sample, the change in average frequency from the first half of the run to the second. The resulting estimate

*Jet Propulsion Laboratory (CIT), 4800 Oak Grove Dr., MS 298-100, Pasadena, CA 91109. E-mail: charles.greenhall@jpl.nasa.gov. The work of this author was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

†Time and Frequency Division, National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80303. E-mail: dhowe@nist.gov.

‡Applied Physics Laboratory, University of Washington, Seattle, WA, and MathSoft, Inc., Seattle, WA 98195. E-mail: dbp@apl.washington.edu. Supported by an ONR ARL Grant (APL, UW) and an NSF Phase II SBIR contract (MathSoft).

$\hat{\sigma}_y(T/2, T)$ often appears to be unrealistically low. An example can be seen in Fig. 1, the results of a test run of a pair of hydrogen masers at JPL's Frequency Standards Laboratory.

Two reasons for the droop at the right end can be given. First, if the increments of the frequency residuals are modeled as Gaussian random variables with mean zero, implying no overall linear frequency drift, then $\hat{\sigma}_y^2(T/2, T)$ is proportional to a chi-square random variable χ_1^2 with one degree of freedom. The distribution of such a random variable is heavily skewed toward values lower than its mean value $\sigma_y^2(T/2)$. This can be seen from Fig. 2, which shows the probability density of the random variable $Q = \log_{10}[\hat{\sigma}_y(T/2, T)/\sigma_y(T/2)]$. The probability that $Q < 0$ is 0.68, more than twice the probability that $Q > 0$, and the left tail is much heavier than the right tail. Second, to prevent frequency drift from masking the long-term fluctuations, it is common practice to remove an estimate of overall linear drift from the data; in this case, $\hat{\sigma}_y^2(T/2, T)$ is likely to be reduced because drift removal tends to match the earlier and later frequencies. After drift removal, $\hat{\sigma}_y^2(T/2, T)$ still has one degree of freedom, so is subject to both effects.

In an effort to reduce these effects on the measurement of $\sigma_y^2(\tau)$ for large τ, the notion of Total variance was developed over the last few years in a sequence of papers [2][3][4][5]. The initial idea is illustrated by Fig. 3, which shows seven cyclic shifts (modulo T) of the uppermost plot, the frequency *sampling function* for estimated Allan variance at $\tau = T/2$: this terminology means that the frequency residuals $y(t)$ are multiplied by this function and summed, giving a linear functional whose square, when properly scaled, is $\hat{\sigma}_y^2(T/2, T)$. This sampling function, which is odd about $T/2$, rejects the even part of $y(t)$. If by chance or design (from the two effects discussed above) it should happen that $y(t)$ tends to be even about $T/2$, then the magnitude of the functional could be much smaller than a practical notion of the size of the long-term frequency variations. In this situation, it makes sense also to apply the even sampling function labeled by $T/4$, which rejects the odd part of $y(t)$. The square root of the sum of the squares of these two linear functionals of $y(t)$ can be expected to have better properties as a measure of long-term stability than the magnitude of either one alone does.

Having admitted one T-cyclic shift of the sampling function, we might as well admit all the others, especially if we want to improve on the fully overlapped unbiased estimator $\hat{\sigma}_y^2(\tau, T)$ of Allan variance [1], henceforth called the *standard* estimator, for averaging times τ that are less than $T/2$. The standard estimator is the scaled mean-square output of the linear functionals generated by all the available time shifts of the $\sigma_y^2(\tau)$ sampling function, $h_\tau(t) = -1$ for $0 < t < \tau$, 1 for $\tau < t < 2\tau$, that fit over the data $y(t)$, $0 \le t \le T$; see (5) below for a formula that applies to discrete-time data. The initial version of Total variance for τ is the scaled mean-square output of the linear functionals generated by all the possible T-cyclic shifts of $h_\tau(t)$. (For $\tau = T/2$, half the sampling functions are redundant because they are the negatives of the other half.)

This version of Total variance enjoyed some success as an estimator of Allan variance with reduced variability and sensitivity to drift removal [2][3], although it seemed to have a problem of increased variability for data dominated by random walk FM. The same estimator can be obtained by fixing the sampling function $h_\tau(t)$ and shifting the *data* cyclically modulo T, or, what is the same, by applying $h_\tau(t)$ as a linear time-invariant filter to an input obtained by extending the original data $y(t)$ periodically with period T. If $y(t)$, $0 \leq t \leq T$, is viewed as a finite piece of an ergodic process, then its T-periodic extension can sometimes be regarded as a substitute for lack of knowledge of the data outside $[0, T]$ [6]. On the other hand, a piece of a random walk, if continued periodically, has a large random discontinuity at the data interfaces 0 and T, untypical of the process as a whole; its effect on the $h_\tau(t)$ functionals cannot be neglected, even for small τ. This problem of mismatched endpoints was solved by the technique of reflecting the data about both endpoints, resulting in a virtual dataset $y^{\#}(t)$ that can be extended to a $2T$-periodic sequence consisting of alternating forward and backward copies of $y(t)$. The lower plot of Fig. 4 shows a portion of $y^{\#}(t)$ about 3 times as long as the original $y(t)$ in the middle section. The current version of Total variance is defined as the scaled mean-square output of the $h_\tau(t)$ filters acting on this new sequence.

The intent of the present paper is to give a precise definition of Total variance and an account of some of its properties. We abbreviate Total variance for τ and T as Totvar(τ, T) or Totvar(τ) (pronounced tōt´-vär). The square root of Total variance is called Total deviation (Totdev). The results given here fall into two classes.

Total variance as a descriptive statistic. Both Totvar(τ, T) and $\hat{\sigma}_y^2(\tau, T)$ are statistics, that is, they are functions only of the data at hand. By a *descriptive* statistic we mean a statistic that has something valuable to say about these data, regardless of any stochastic model that might be fitted or any assumptions about how the data might have evolved outside the interval of observation. Simple examples are sample mean and sample variance. In Section 3 we show that Total variance can be used to carry out an analysis of variance, an exact decomposition of the sample variance s_y^2 of the frequency residuals $y_n = y(n\tau_0)$, where τ_0 is the sample period. In particular, Totvar$\left(2^j\tau_0\right)$ (when rescaled) can be regarded as the portion of s_y^2 to be associated with the octave frequency band $2^{-j-2}/\tau_0 < \nu < 2^{-j-1}/\tau_0$. Thus, after evaluating Totvar(τ) for $\tau = \tau_0, 2\tau_0, \ldots, 2^j\tau_0$, one can tell how much of the sample variance is yet unaccounted for, and associate the low-frequency band $0 < \nu < 2^{-j-2}/\tau_0$ with the remainder. Analysis of sample variance is a central theme in statistics; an exact decomposition is highly desirable because it accounts for all the observed variance in the data. The periodogram does such, as do most spectrum estimators and other decompositions, but the standard estimator $\hat{\sigma}_y^2(\tau, T)$ of Allan variance does not [7].

Total variance as an estimator of Allan variance. Presented below are results for the mean and variance of $\mathrm{Totvar}(\tau, T)$ in the presence of three power-law FM noises: white FM, flicker FM, and random walk FM. For white FM we find that $\mathrm{Totvar}(\tau, T)$ is an unbiased estimator of $\sigma_y^2(\tau)$ for $0 < \tau \leq T$. The bias of $\mathrm{Totvar}(\tau, T)$ at $\tau = T/2$ is -24% for flicker FM and -37.5% for random walk FM, and has the simple form $a\tau/T$ for $0 < \tau \leq T/2$. (These biases apply to σ^2, not σ.) The equivalent degrees of freedom (edf; see (22) below) of $\mathrm{Totvar}(\tau, T)$ is always greater than that of $\hat{\sigma}_y^2(\tau, T)$; for $\tau = T/2$ in particular, the edf of Total variance is 3 for white FM, 2.1 for flicker FM, and 1.5 for random walk FM, while the edf of the standard estimator is 1. Moreover, the edf of $\mathrm{Totvar}(\tau, T)$, $0 < \tau \leq T/2$, can be well approximated by first-degree polynomials in T/τ for each noise type. The mean square error of $\mathrm{Totvar}(\tau, T)$ is less than that of $\hat{\sigma}_y^2(\tau, T)$, even though the former is biased and the latter is not. Confidence intervals for $\sigma_y^2(\tau)$ based on a chi-square assumption for $\mathrm{Totvar}(\tau, T)$ can easily be constructed; these will be tighter than those based on $\hat{\sigma}_y^2(\tau, T)$, and there is evidence that such confidence intervals are conservative.

In summary, Total variance is presented as a tool for squeezing a modest amount of extra information about long-term stability from a dataset of clock residuals, information that is often obscured by the standard Allan variance estimator for τ at or near $T/2$. Analyzing frequency stability accurately in long term has been problematic even for experienced users. The properties of Total variance presented here suggest that it uses the available data more efficiently than the standard estimator for long-term characterizations. Confident of these properties, the authors expect to see wider usage of this tool.

2 Definition of Total variance

The purpose of this section is to give a precise definition of $\mathrm{Totvar}(\tau, T)$ for an N_x-point time-residual record with sample period τ_0. In the following description, the indices m, n, and N_x are related to time by $\tau = m\tau_0$, $t = t_0 + n\tau_0$, and $T = (N_x - 1)\tau_0$, where t_0 is the time origin and without loss may be made equal to 0.

We start with time-residual data $x_1, \ldots, x_{N_x}$, with normalized frequency residuals $y_n = (x_{n+1} - x_n)/\tau_0$, $1 \le n \le N_y = N_x - 1$. Extend the sequence y_n to a new, longer virtual sequence $y_n^{\#}$ by reflection as follows: for $1 \le n \le N_y$ let $y_n^{\#} = y_n$; for $1 \le j \le N_y - 1$ let

$$y_{1-j}^{\#} = y_j, \quad y_{N_y+j}^{\#} = y_{N_y+1-j}. \tag{1}$$

An equivalent operation can be performed on the original time-residual sequence x_n to produce an extended virtual sequence $x_n^{\#}$ as follows: for $1 \le n \le N_x$ let $x_n^{\#} = x_n$; for $1 \le j \le N_x - 2$ let

$$x_{1-j}^{\#} = 2x_1 - x_{1+j}, \quad x_{N_x+j}^{\#} = 2x_{N_x} - x_{N_x-j}. \tag{2}$$

This operation, illustrated in Fig. 4 by the record used for Fig. 1, is called extension by reflection about both endpoints. The result of this extension is a virtual data sequence $x_n^{\#}$, $3 - N_x \le n \le 2N_x - 2$, having length $3N_x - 4$ and satisfying $y_n^{\#} = \left(x_{n+1}^{\#} - x_n^{\#}\right)/\tau_0$, $3 - N_x \le n \le 2N_x - 3$.

We now define

$$\mathrm{Totvar}(m, N_x, \tau_0) = \frac{1}{2(m\tau_0)^2 (N_x - 2)} \sum_{n=2}^{N_x-1} \left(x_{n-m}^{\#} - 2x_n^{\#} + x_{n+m}^{\#}\right)^2, \tag{3}$$

for $1 \le m \le N_x - 1$. Note that τ is allowed to go to $(N_x - 1)\tau_0$ instead of the usual limit of $\lfloor (N_x - 1)/2 \rfloor \tau_0$. Total variance can also be represented in terms of extended fractional frequency residual averages by

$$\mathrm{Totvar}(m, N_y + 1, \tau_0) = \frac{1}{2(N_y - 1)} \sum_{n=2}^{N_y} \left[\bar{y}_n^{\#}(m) - \bar{y}_{n-m}^{\#}(m)\right]^2, \tag{4}$$

where $\bar{y}_n^{\#}(m) = \left(x_{n+m}^{\#} - x_n^{\#}\right)/(m\tau_0)$.

The notations $\mathrm{Totvar}(\tau, T)$ and $\mathrm{Totvar}(\tau)$ are to be regarded as abbreviations for $\mathrm{Totvar}(m, N_x, \tau_0)$.

2.1 Remarks

- For comparison, the standard Allan variance estimator, which we have been abbreviating as $\hat{\sigma}_y^2(\tau, T)$, is actually given by

$$\hat{\sigma}_y^2(m, N_x, \tau_0) = \frac{1}{2(N_x - 2m)} \sum_{n=1}^{N_x-2m} \left[\bar{y}_{n+m}(m) - \bar{y}_n(m)\right]^2, \tag{5}$$

  where $1 \le m < N_x/2$, $\bar{y}_n(m) = (x_{n+m} - x_n)/(m\tau_0)$.

- Total variance, like Allan variance and its conventional estimators, is invariant to an overall shift in phase and frequency; that is, if a first-degree polynomial $c_0 + c_1 n$ is added to the original dataset x_n, then Total variance does not change.

- It is possible to program Total variance without creating an extended data array in memory [5, Eq. (9)].
- To simplify the scaling of Total variance relative to the decomposition of sample variance that it generates (see (17)–(19) below), one might wish to use $N_x - 1$ in place of the denominator $N_x - 2$ in (3), and N_y in place of the denominator $N_y - 1$ in (4).
- In Section 3.3, the definition of Total variance is extended to arbitrarily large m.

3 Analysis of variances

3.1 Multiresolution scheme

The variance decomposition properties of Allan variance and Total variance can be derived from an overlapped Haar wavelet transform [8]. The scheme consists of a ladder of linear time-invariant filters (Fig. 5), which, acting on an input sequence y_n with sample period τ_0, decomposes the original frequency range $0 < \nu < 2^{-1}/\tau_0$ into successively lower octave bands; each stage leaves a smoothed version of the input for further analysis. The ladder is built from two simple filters: a lowpass filter G_0 with impulse response $[g_0, g_1] = \frac{1}{2}[1, 1]$, and a highpass filter H_0 with impulse response $[h_0, h_1] = \frac{1}{2}[1, -1]$. The corresponding transfer functions,

$$G_0(\nu) = e^{-i\pi\nu\tau_0}\cos(\pi\nu\tau_0), \quad H_0(\nu) = ie^{-i\pi\nu\tau_0}\sin(\pi\nu\tau_0),$$

satisfy

$$|G_0(\nu)|^2 + |H_0(\nu)|^2 = 1. \tag{6}$$

Let F be a filter with impulse response f_n and transfer function $F(\nu)$. For any positive integer r, define the r-*upsampled* version of F to have an impulse response $f_j^{(r)}$ such that $f_{nr}^{(r)} = f_n$ for all n, and $f_j^{(r)} = 0$ otherwise; in other words, $r - 1$ zeros are inserted between successive terms. The transfer function of the upsampled filter is $F(r\nu)$.

For $j = 1, 2, \ldots$ let G_j and H_j be the 2^j-upsampled versions of G_0 and H_0. These filters are applied to y_n according to the scheme shown in Fig. 5. Its jth stage has output sequences $v_{j,n} = A_j y_n$ and $w_{j,n} = B_j y_n$, where $A_0 = G_0$, $B_0 = H_0$, and

$$A_j = G_j G_{j-1} \cdots G_0, \quad B_j = H_j G_{j-1} \cdots G_0 = H_j A_{j-1} \tag{7}$$

for $j \geq 1$. One can show by induction that A_j is a moving-average filter with impulse response $2^{-j-1}[1, \ldots, 1]$ (2^{j+1} ones), a lowpass filter with bandwidth $2^{-j-2}/\tau_0$. Then, by (7), B_j has impulse response $2^{-j-1}[1, \ldots, 1, -1, \ldots, -1]$ (2^j ones, then 2^j minus-ones), which is $2^{-1/2}$ times the filter associated with $\sigma_y^2(2^j \tau_0)$. This filter is an approximate bandpass filter for the octave band $2^{-j-2}/\tau_0 < \nu < 2^{-j-1}/\tau_0$ [9]. The squared frequency responses of A_j and B_j are given by

$$|A_j(\nu)|^2 = \frac{\sin^2(2^{j+1}\pi\nu\tau_0)}{4^{j+1}\sin^2(\pi\nu\tau_0)}, \quad |B_j(\nu)|^2 = \frac{\sin^4(2^j\pi\nu\tau_0)}{4^j\sin^2(\pi\nu\tau_0)}.$$

As we have seen, the approximate passbands of the filters $B_0, \ldots, B_J, A_J$ form a partition of the original frequency domain $0 < \nu < 2^{-1}/\tau_0$. The variance decompositions discussed below follow from a precise counterpart of this statement for the squared frequency responses, which satisfy the frequency-domain decomposition

$$|A_J(\nu)|^2 + \sum_{j=0}^{J} |B_j(\nu)|^2 = 1. \tag{8}$$

This equation can be proved by induction on J from (6) and (7) (or from the identity $\sin^4 x = \sin^2 x - \frac{1}{4}\sin^2 2x$). If $\nu\tau_0$ is not an integer, then $|A_J(\nu)|^2 \to 0$ as $J \to \infty$, and it follows from (8) that

$$\sum_{j=0}^{\infty} |B_j(\nu)|^2 = 1, \tag{9}$$

$$|A_J(\nu)|^2 = \sum_{j=J+1}^{\infty} |B_j(\nu)|^2. \tag{10}$$

Eq. (9) says that the squared frequency responses associated with Allan variance for $\tau = 2^j\tau_0$ sum to 2, except at zero frequency and its aliases.

3.2 Ensemble variance

Before deriving the sample variance decomposition property of Total variance, it is useful to see how the analogous property of Allan variance follows from the frequency-domain decompositions. Let y_n be a stationary random process with variance σ_y^2 and one-sided spectral density $S_y(\nu)$. Multiplying (8)–(10) by $S_y(\nu)$ and integrating over $0 < \nu < 2^{-1}/\tau_0$, we obtain

$$\sigma_y^2 = \sigma_{v_J}^2 + \frac{1}{2}\sum_{j=0}^{J} \sigma_y^2(2^j\tau_0) = \frac{1}{2}\sum_{j=0}^{\infty} \sigma_y^2(2^j\tau_0), \tag{11}$$

$$\sigma_{v_J}^2 = \frac{1}{2}\sum_{j=J+1}^{\infty} \sigma_y^2(2^j\tau_0). \tag{12}$$

If y_n has stationary first increments but is not stationary, like flicker FM and random-walk FM, then $w_{j,n}$ is stationary, $v_{j,n}$ is not stationary, and the frequency-domain integrals giving σ_y^2 and $\sigma_{v_J}^2$ are infinite, as are the infinite series in (11) and (12).

3.3 Sample variance

From the random-process setting we return to the consideration of a finite data sequence $y_1, \ldots, y_N$ with sample period τ_0. Before invoking the extension procedure of Total variance, let us consider temporarily a simpler periodic extension with period N, that is, we agree that $y_{n+N} = y_n$ for all integers n. The sample mean m_y and sample variance s_y^2 of y_n are conveniently expressed in terms of its discrete Fourier transform (DFT), given by

$$Y_k = \sum_{n=1}^{N} y_n e^{-i2\pi kn/N},$$

and also indicated by the notation $y_n \longleftrightarrow Y_k$. We have $m_y = Y_0/N$, and

$$s_y^2 = \frac{1}{N} \sum_{n=1}^{N} y_n^2 - m_y^2 = \frac{1}{N^2} \sum_{k=1}^{N-1} |Y_k|^2 \tag{13}$$

by Parseval's theorem in the DFT setting.

Let F be a filter with finite impulse response f_n and transfer function $F(\nu)$. Define $f_n(N)$ to be the sum of f_j over all j such that $\bmod(j, N) = n$. Some facts about the periodic sequence $f_n(N)$ must now be set down. First, if y_n is N-periodic, then so is Fy_n, and we have

$$Fy_n = \sum_{j=1}^{N} f_j(N)\, y_{n-j},$$

which expresses a circular convolution; the summation can be taken over any period. Second, $f_n(N) \longleftrightarrow F(\nu_k)$, where $\nu_k = k/(N\tau_0)$. Therefore, $Fy_n \longleftrightarrow F(\nu_k)\, Y_k$.

Let the input to our multiresolution scheme be an N-periodic sequence y_n. Then all the output sequences are periodic, and from the previous remarks it follows that $v_{j,n} \longleftrightarrow A_j(\nu_k)\, Y_k$, $w_{j,n} \longleftrightarrow B_j(\nu_k)\, Y_k$, and

$$s_{v_j}^2 = \frac{1}{N^2} \sum_{k=1}^{N-1} |A_j(\nu_k)|^2 |Y_k|^2, \quad s_{w_j}^2 = \frac{1}{N^2} \sum_{k=1}^{N-1} |B_j(\nu_k)|^2 |Y_k|^2. \tag{14}$$

Combining (14) and (13) with the frequency-domain partitions (8)–(10) yields the analogs of (11)–

(12) for sample variances:

$$s_y^2 = s_{v_J}^2 + \sum_{j=0}^{J} s_{w_j}^2 = \sum_{j=0}^{\infty} s_{w_j}^2, \tag{15}$$

$$s_{v_J}^2 = \sum_{j=J+1}^{\infty} s_{w_j}^2. \tag{16}$$

Observe also that $m_{v_j} = m_y$, $m_{w_j} = 0$ because $A_j(0) = 1$, $B_j(0) = 0$.

Returning to the Total variance setting of Section 2, we set $N = 2N_y$ and replace $[y_1, \ldots, y_N]$ by $y^{\#} = \left[y_1, \ldots, y_{N_y}, y_{N_y}, \ldots, y_1\right]$, which, if extended to a sequence with period $2N_y$, agrees with the definition given in Section 2. Because of the symmetry of the $y_n^{\#}$ about their midpoint and the symmetry of the filters A_j and B_j, the terms in the various sums of squares occur in pairs. As a result, $s_{y^{\#}}^2 = s_y^2$ (the sample variance of $\left[y_1, \ldots, y_{N_y}\right]$), $s_{v_j}^2$ is the sample variance of the 2^{j+1}-moving average of $y_n^{\#}$, and

$$s_{w_j}^2 = \frac{N_y - 1}{2N_y} \operatorname{Totvar}\left(2^j \tau_0\right). \tag{17}$$

The expression (4) is contrived so that only $N_y - 1$ distinct nonzero terms are used. Although Totvar was previously defined only for $m < N_y$, (4) retains meaning for all m if $y_n^{\#}$ is extended periodically as far as needed.

To simplify the equations of analysis of variance involving Totvar, it is convenient to define a "Remainder variance," Remvar $(m\tau_0)$, as $2N_y(N_y - 1)^{-1}$ times the sample variance of $2N_y$ successive values of the moving m-averages of the $2N_y$-periodic sequence $y_n^{\#}$. Then

$$s_y^2 = \frac{N_y - 1}{2N_y} \operatorname{Remvar}(\tau_0), \tag{18}$$

$$s_{v_j}^2 = \frac{N_y - 1}{2N_y} \operatorname{Remvar}\left(2^{j+1} \tau_0\right). \tag{19}$$

The square root of Remainder variance is called Remainder deviation (Remdev).

In this setting, the variance decompositions (15)–(16) become

$$\operatorname{Remvar}(\tau_0) = \operatorname{Remvar}\left(2^{J+1}\tau_0\right) + \sum_{j=0}^{J} \operatorname{Totvar}\left(2^j \tau_0\right), \tag{20}$$

$$\operatorname{Remvar}\left(2^J \tau_0\right) = \sum_{j=J}^{\infty} \operatorname{Totvar}\left(2^j \tau_0\right), \tag{21}$$

for $J \geq 0$. In other words, the rescaled sample variance of y_n is the sum of all the Totvar $\left(2^j \tau_0\right)$, and Remvar $\left(2^{J+1}\tau_0\right)$ indicates how much of the sample variance has not yet been accounted for by Totvar $\left(2^j \tau_0\right)$, $j \leq J$.

3.3.1 Remarks

- Total variance is the first "modern" estimator of Allan variance to mimic its ensemble variance decomposition properties (11)–(12); moreover, the sample variance decompositions (20)–(21) apply to any finite dataset.

- Because higher order Daubechies wavelet filters [10] also satisfy (6), the above development extends easily to higher order wavelet variances (Allan variance is essentially twice the Haar wavelet variance). These higher order wavelet variances are suitable substitutes for some of the variations on Allan variance that have been proposed and studied in the literature (modified Allan variance, for example). For details, see [11].

- As mentioned above, we can compute $\text{Totvar}\,(\tau, T)$ for arbitrarily large τ without taking more data, but $\text{Totvar}\,(\tau, T)$ for $\tau > T$ ought not to be regarded as an estimator of $\sigma_y^2\,(\tau)$. In particular, if $N_y = 2^K$ then $\text{Remvar}\left(2^j \tau_0\right)$ and $\text{Totvar}\left(2^j \tau_0\right)$ vanish for $j \geq K + 1$.

- See [7] for a discussion of analysis of sample variance by the non-overlapped estimator of Allan variance.

3.3.2 Example

Fig. 1 shows $\hat{\sigma}_y\,(\tau, T)$, $\text{Totdev}(\tau, T)$, and $\text{Remdev}(\tau, T)$ for the hydrogen-maser record shown in Fig. 4, with parameters $N_x = 1727$, $\tau_0 = 1024.1\,\text{s}$, $T = 1.77 \times 10^6\,\text{s}$. The averaging times include $2^j \tau_0,\ 0 \leq j \leq 11$, and also include $T/2 = 8.84 \times 10^5\,\text{s}$. Relative frequency drift, -1.81×10^{-15} per day, was estimated from the data by the "4-point w" method [12] and removed. As a result, $\hat{\sigma}_y\,(T/2, T)$ becomes severely depressed, a common consequence of drift removal. (It would have been identically zero if the "3-point x" drift estimate [13] had been used.) On the other hand, $\text{Totdev}(\tau)$ shows no depression until τ exceeds $T/2$. The flatness of Remdev at the lower τ means that the first several values of Totvar do not contribute much to the sample variance of the frequency record (essentially $\frac{1}{2}\,\text{Remvar}\,(\tau_0)$). At $\tau = 2^{10}\tau_0$, Remdev and Totdev are almost equal, which indicates that $\text{Remvar}\left(2^{10}\tau_0\right)$ is the last significant component of the sample variance.

The error bars, which are confidence intervals for $\sigma_y\,(\tau)$ based upon $\text{Totdev}(\tau, T)$, are discussed below.

4 Totvar as Allan variance estimator

Although Total variance can stand on its own as a descriptive statistic that performs an analysis of variance on a dataset, its usefulness for time and frequency measurement is based mainly on its statistical properties as an estimator of Allan variance under assumptions about the underlying random noise processes. Because we are interested mostly in long-term frequency stability statistics, and for mathematical convenience, we treat only the power-law FM noise processes that are likely to dominate long-term measurements: white FM, flicker FM, and random walk FM (WHFM, FLFM, RWFM). With these assumptions, the properties of $\mathrm{Totvar}\,(m, N_y + 1, \tau_0)$ depend mainly on the ratio $m/N_y = \tau/T$ for large m and N_y; it is convenient, then, to approximate $\mathrm{Totvar}\,(\tau, T)$ by a continuous-time analog in which sums involving x_n are replaced by integrals involving $x\,(t)$, modeled by a power-law process with spectral density proportional to $\nu^{\alpha-2}$ for $0 < \nu < \infty$ (no high-frequency cutoff) and $\alpha = 0, -1, -2$ [5]. The theoretical computations also assume that there is no underlying linear frequency drift, i.e., that the second increments of x have mean zero, and that these increments are Gaussian.

4.1 Mean and variance

Although Total variance is most conveniently expressed as a function of the extended record $x_n^{\#}$ or $y_n^{\#}$, each term of (3) can also be expressed as the square of a linear functional of the original data sequence x_n or y_n. These functionals, though complicated by the foldover implicit in the extension by reflection, are still second-order functionals of the x_n, that is, they are invariant to time and frequency shifts. (See [5] for formulas and illustrations of the sampling functions.) Therefore, it is possible to compute the mean and variance of the quadratic functional that constitutes Totvar in the presence of phase noise with stationary, mean-zero, Gaussian second increments. These computations were performed by manipulations on the generalized autocovariances of the three FM noise processes [14].

The mean $E\,[\mathrm{Totvar}\,(\tau, T)]$ is compared to $\sigma_y^2\,(\tau)$; the variance is most conveniently communicated through the equivalent degrees of freedom (edf), defined for a random variable V by

$$\mathrm{edf}\,(V) = \frac{2\,(EV)^2}{\mathrm{var}\,(V)}. \tag{22}$$

The results can be expressed by the formulas

$$r := \frac{E\left[\mathrm{Totvar}\left(\tau, T\right)\right]}{\sigma_y^2\left(\tau\right)} = 1 - a\frac{\tau}{T}, \quad 0 < \tau \leq \frac{T}{2}, \tag{23}$$

$$\nu := \mathrm{edf}[\mathrm{Totvar}\left(\tau, T\right)] \approx b\frac{T}{\tau} - c, \quad 0 < \tau \leq \frac{T}{2}, \tag{24}$$

where a, b, and c are given in Table 1. The values of r and ν for the important longest-term case $\tau = T/2$ are also tabulated. The edf formula (24) is empirical, with an observed error below 1.2% of numerically computed exact values; the tabulated values of $\nu\left(T/2\right)$ are the exact ones. As is obvious from their form, a and b were derived from theory; in particular, b is the limiting value of $(\tau/T)\mathrm{edf}\left[\hat{\sigma}_y^2\left(\tau, T\right)\right]$ for large T/τ [15]. The only coefficient that had to be chosen empirically was c. These results were checked by simulations of $\mathrm{Totvar}\left(m, N_x, \tau_0\right)$, with $N_x = 101$. The simplicity, accuracy, and range of applicability of (24) are striking in view of existing approximations for the edf of $\hat{\sigma}_y^2\left(\tau, T\right)$ [1][15][16]. Although Total variance is an estimator of greater complexity, some of its statistical properties are simpler.

Fig. 6 compares $\mathrm{Totvar}\left(\tau, T\right)$ to the standard unbiased Allan variance estimator $\hat{\sigma}_y^2\left(\tau, T\right)$ in two different ways. The upper plot shows the ratio of the edf of the two estimators for the three FM noises; the lower plot shows the ratio of their root-mean-square errors ($\sqrt{\mathrm{bias}^2 + \mathrm{variance}}$) as estimators of $\sigma_y^2\left(\tau\right)$.

4.1.1 Remarks

- Because of the continuous-time analog used for the theoretical calculations, (24) should be used only if $\tau \geq 8\tau_0$ for white FM, $3\tau_0$ for flicker FM.

- The simple, exact form (23) for the mean of Totvar can be interpreted as a scaling property of power-law noise. It turns out this way because the shapes of the sampling functions for Total variance [5] depend only on τ when $\tau \leq T/2$. For $T/2 < \tau \leq T$ the sampling function shapes depend also on T; yet, it is noteworthy that (23) persists all the way to T, but only for white FM and random walk FM.

- For white FM, $\mathrm{Totvar}\left(\tau, T\right)$ is an unbiased estimator of $\sigma_y^2\left(\tau\right)$ for $\tau \leq T$. This fact appeared as an outcome of algebraic manipulations; unfortunately, the authors cannot give a simple reason why it is so. The edf result (24) for white FM, though obtained numerically, seems to be exact for $\tau \leq T/2$ and $\tau = T$ with $c = 0$. No calculations of edf were performed for $T/2 < \tau < T$.

4.2 Confidence intervals

In the tradition of time and frequency statistics, it is customary to derive confidence intervals for frequency stability on the basis of the assumption that the probability distribution of a frequency stability estimator V, when scaled appropriately, follows the chi-squared distribution with the same edf as V [1]. Fix τ, and write $V = \text{Totvar}(\tau, T)$, $\sigma^2 = \sigma_y^2(\tau)$. Let $r = E(V)/\sigma^2$, $\nu = \text{edf}(V)$, given by (23)–(24) in the presence of one of the three FM noises. Then the random variable

$$X = \frac{\nu V}{r\sigma^2} \tag{25}$$

has the same mean and edf, namely ν, as a χ_ν^2 variable does. Presume for the moment that X has a χ_ν^2 distribution. For $0 \le p_1 < p_2 < 1$ let ξ_1 and ξ_2 be the corresponding levels of this distribution. (A simple approximation algorithm for χ_ν^2 levels can be found in [17].) Then $\xi_1 < X < \xi_2$ with probability $p = p_2 - p_1$; a rearrangement gives the confidence statement that

$$\frac{\nu V}{r\xi_2} < \sigma^2 < \frac{\nu V}{r\xi_1} \tag{26}$$

with probability p. Observe that the bias of V has been allowed to push the confidence interval upwards (when $r < 1$).

The error bars in Fig. 1, shifted horizontally for visibility, are 90% confidence intervals for $\sigma_y(\tau)$ as computed by this method ($p_1 = 0.05$, $p_2 = 0.95$) under the assumption of flicker FM and random walk FM noise models. Both sets of error bars suggest the hypothesis that random walk FM is the dominant noise type for $10^5\text{s} < \tau < 10^6\,\text{s}$, although a flicker FM hypothesis is not ruled out. A longer test run ($T = 4.23 \times 10^6$ s) of the same pair of standards supports the random walk hypothesis, with $\hat{\sigma}_y(\tau, T)$ increasing like $\tau^{1/2}$. On the other hand, the longer run has a sharp frequency step of about 4×10^{-14}, untypical of the shorter run (Fig. 4), so that the authors hesitate to declare a successful characterization.

We note that (23)–(24) have not been shown to be accurate when estimated frequency drift is removed, as was done for Fig. 1. The authors have not carried out the required theoretical computations, which depend on the method of drift estimation and are more intricate than before. It seems clear, though, that the effect of drift removal on Total variance is less than its effect on conventional estimators of $\sigma_y^2(\tau)$, which tend to be severely depressed for τ near $T/2$ [3].

The χ_ν^2 assumption for Total variance has been investigated, for $\tau = T/2$ only, by simulation of the three FM noise types [5]. The empirical distributions of X as defined by (25) were observed to have heavier left tails than those of the corresponding χ_ν^2 distributions. If this turns out to be true in general, it means that the *upper* ends of confidence intervals (26) based on the χ^2_ν distribution are pessimistic. For now, use of the χ^2_ν distribution for this purpose seems to be a conservative policy.

References

[1] D. A. Howe, D. W. Allan, and J. A. Barnes, "Properties of Signal Sources and Measurement Methods," *Proc. 35th Annual Symposium on Frequency Control*, pp. 1–47, 1981.

[2] D. A. Howe, "An Extension of the Allan Variance with Increased Confidence at Long Term," *Proc. IEEE International Frequency Control Symposium*, pp. 321–329, 1995.

[3] D. A. Howe and K. J. Lainson, "Effect of Drift on TOTALDEV," *Proc. IEEE International Frequency Control Symposium*, pp. 883–889, 1996.

[4] D. A. Howe, "Methods of Improving the Estimation of Long-term Frequency Variance," *Proc. European Frequency and Time Forum*, pp. 91–99, 1997.

[5] D. A. Howe and C. A. Greenhall, "Total variance: a progress report on a new frequency stability characterization," *Proc. 29th Ann. PTTI Systems and Applications Meeting*, pp. 39–48, 1997.

[6] D. A. Howe, "Circular Representation of Infinitely Extended Sequences," *Proc. IEEE International Frequency Control Symposium*, pp. 337–345, 1995.

[7] D. B. Percival and D. A. Howe, "Total variance as an exact analysis of the sample variance," *Proc. 29th Ann. PTTI Systems and Applications Meeting*, pp. 97–105, 1997.

[8] D. B. Percival and H. O. Mofjeld, "Analysis of subtidal coastal sea level fluctuations using wavelets," *J. American Statistical Assoc.*, vol. 92, pp. 868–880, 1997.

[9] D. B. Percival, "Characterization of frequency stability: frequency-domain estimation of stability measures," *Proc. IEEE*, vol. 79, pp. 961–972, 1991.

[10] I. Daubechies, *Ten Lectures on Wavelets*, Philadelphia: SIAM, 1992.

[11] D. A. Howe and D. B. Percival, "Wavelet Variance, Allan Variance, and Leakage," *IEEE Trans. Instrum. Meas.*, vol. IM-44, pp. 94–97, 1995.

[12] C. A. Greenhall, "A frequency-drift estimator and its removal from modified Allan variance," *Proc. IEEE International Frequency Control Symposium*, pp. 428–432, 1997.

[13] M. A. Weiss and C. Hackman, "Confidence on the three-point estimator of frequency drift," *Proc. 24th Ann. PTTI Applications and Planning Meeting*, pp. 451–455, 1992.

[14] C. A. Greenhall, "A Structure Function Representation Theorem with Applications to Frequency Stability Estimation," *IEEE Trans. Instrum. Meas.*, vol. IM-32, pp. 364–370, 1983.

[15] C. A. Greenhall, "Recipes for degrees of freedom of frequency stability estimators," *IEEE Trans. Instrum. Meas.*, vol. 40, pp. 994–999, 1991.

[16] K. Yoshimura, "Degrees of freedom of the estimate of the two-sample variance in the continuous-sampling method," *IEEE Trans. Instrum. Meas.*, vol. 38, pp. 1044–1049, 1989.

[17] M. A. Weiss and C. A. Greenhall, "A simple algorithm for approximating confidence on the modified Allan variance and the time variance," *Proc. 28th Ann. PTTI Systems and Applications Meeting*, pp. 215–224, 1996.

Table 1: Coefficients for computing the normalized mean r and edf ν of Total variance for FM noises. Tabulated also are the exact quantities for $\tau = T/2$.

Noise	a	b	c	$r\,(T/2)$	$\nu\,(T/2)$
WHFM	0	3/2	0.000	1	3.000
FLFM	$(3\ln 2)^{-1}$	$24\,(\ln 2)^2\,\pi^{-2}$	0.222	0.760	2.097
RWFM	3/4	140/151	0.358	5/8	1.514

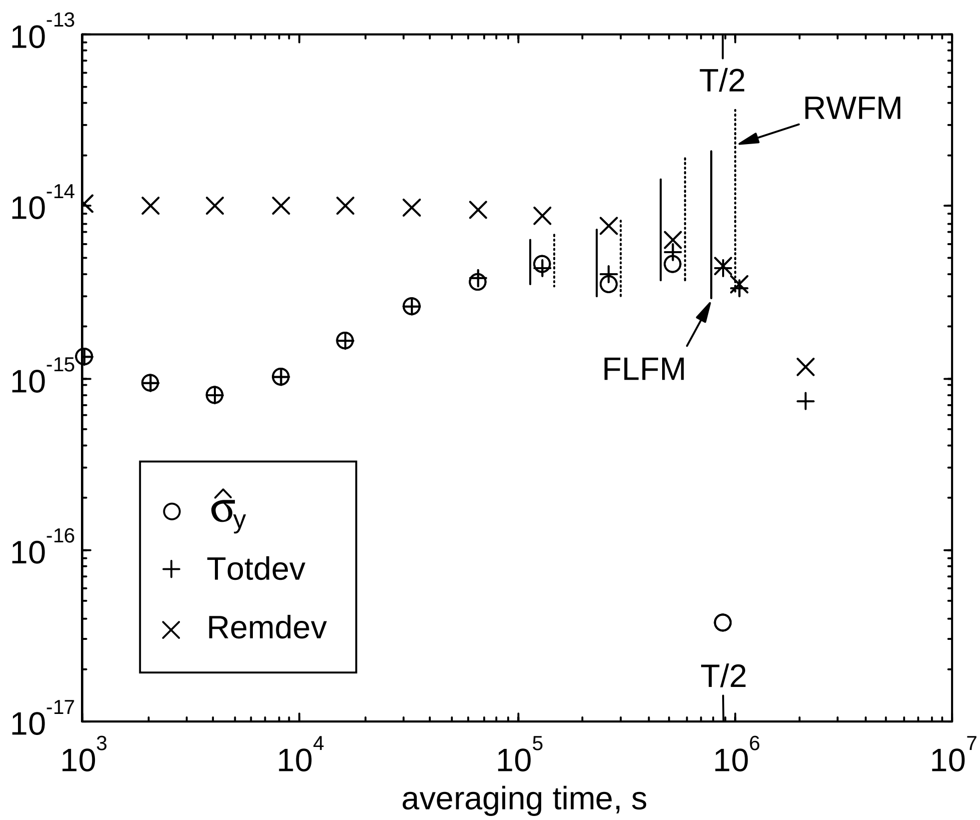

Figure 1: Sigma-tau plot of $\hat{\sigma}_y$ (standard estimate of Allan deviation), Total deviation, and Remainder deviation for a pair of hydrogen masers. The error bars are 90% confidence intervals for Allan deviation based upon Total deviation.

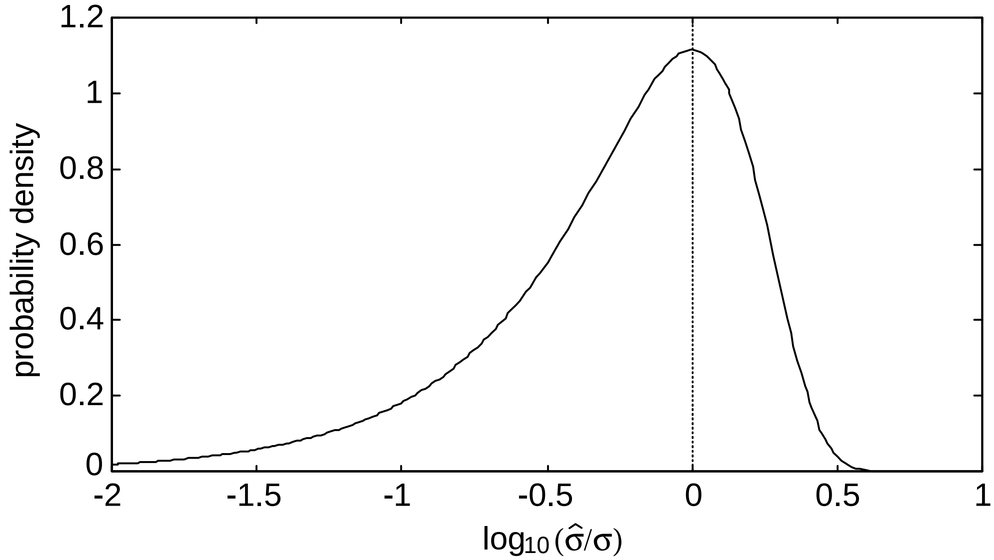

Figure 2: Probability density function of the logarithm (base 10) of the normalized standard estimator $\hat{\sigma}_y\,(T/2, T)$, whose square has one degree of freedom.

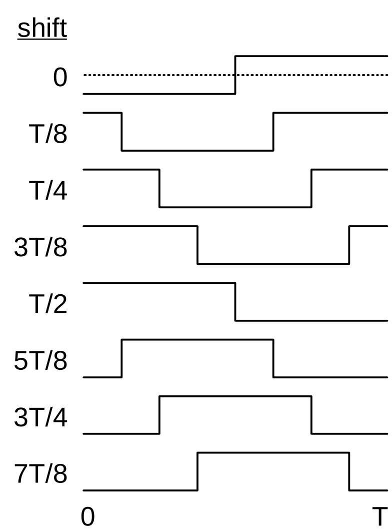

Figure 3: Illustrating the first concept of Total variance: cyclic shifts of the frequency sampling function (uppermost plot) for $\hat{\sigma}^2_y\,(T/2, T)$.

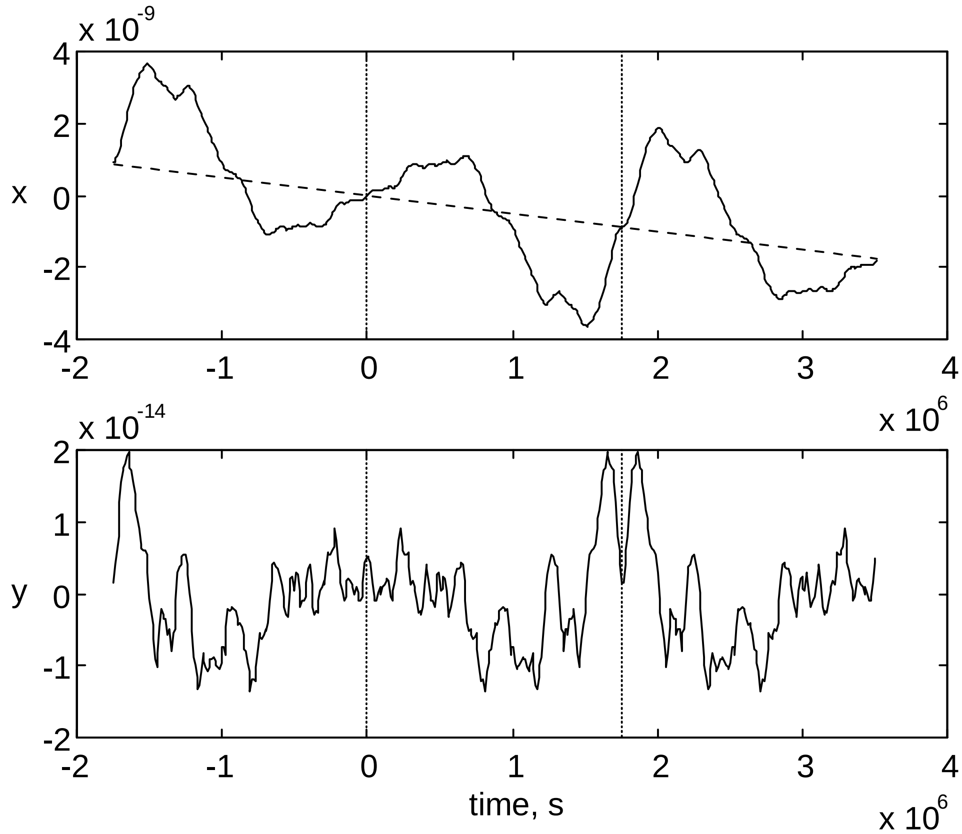

Figure 4: Extension by reflection of time and normalized frequency residuals for computation of Total variance. The original data are between the dotted lines.

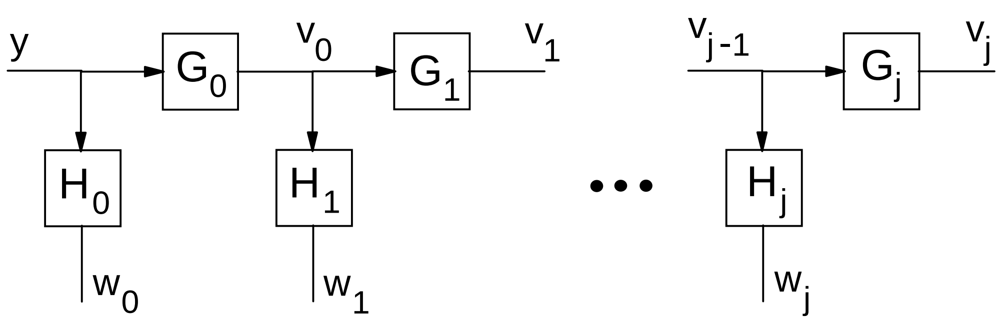

Figure 5: Multiresolution scheme that leads to variance decompositions by Allan variance and Total variance.

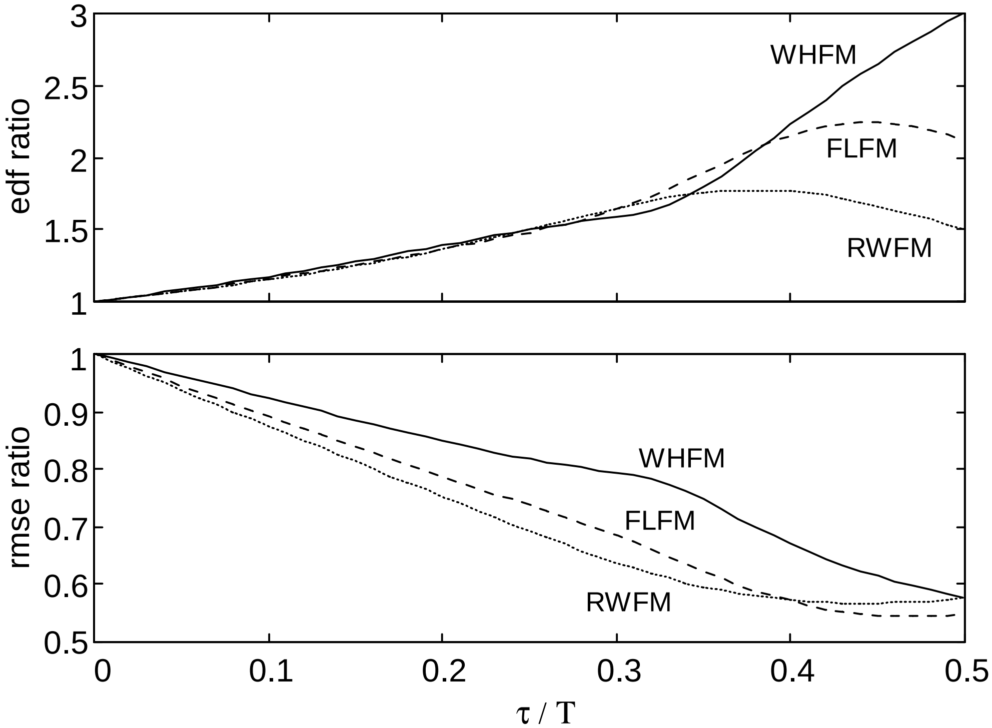

Figure 6: Ratios of equivalent degrees of freedom and root-mean-square error for Total variance over the standard estimator of Allan variance.